This is the accepted manuscript of the article, which has
been published in Journal of Voice, 2019, 33(4), 501-509.
https://doi.org/10.1016/j.jvoice.2018.01.012

Emotion recognition from singing voices using contemporary commercial music and classical styles

[1)]Tua Hakanpää, [1, 2)]Teija Waaramaa, [1)]Anne-Maria Laukkanen

1) Speech and Voice Research Laboratory, Faculty of Education, University of Tampere, Tampere, Finland
2) Faculty of Communication Sciences, University of Tampere, Tampere, Finland

Address for correspondence

Tua Hakanpää, MA

Vanhan-Mankkaan tie 16 e 14

02180 Espoo

Hakanpaa.tua.s@student.uta.fi

+358 405 681172

Abstract

Objectives: This study examines the recognition of emotion in contemporary commercial music (CCM) and classical styles of singing. This information may be useful in improving the training of interpretation in singing.

Study design: Experimental comparative study

Methods: Thirteen singers (11 female, 2 male) with a minimum of 3 years' professional-level singing studies (in CCM or classical technique or both) participated. They sang at three pitches (females a, e1, a1, males one octave lower) expressing anger, sadness, joy, tenderness, and a neutral state. Twenty-nine listeners listened to 312 short (0.63–4.8 s) voice samples, 135 of which were sung using a classical singing technique and 165 of which were sung in a CCM style. The listeners were asked which emotion they heard. Activity and valence were derived from the chosen emotions.

Results: The percentage of correct recognitions out of all the answers in the listening test (N=9,048) was 30.2%. The recognition percentage for the CCM-style singing technique was higher (34.5%) than for the classical-style technique (24.5%). Valence and activation were better perceived than the emotions themselves, and activity was better recognized than valence. A higher pitch was more likely to be perceived as joy or anger, and a lower pitch as sorrow. Both valence and activation were better recognized in the female CCM samples than in the other samples.

Conclusions: There are statistically significant differences in the recognition of emotions between classical and CCM styles of singing. Furthermore, in the singing voice, pitch affects the perception of emotions, and valence and activity are more easily recognized than emotions.

Keywords: voice quality, emotion expression, perception, song genre

1.

Introduction

Singers need to express emotion vocally with great passion, but with sufficient control that the audience can identify with the portrayal of emotion and at the same time enjoy the brilliance of the musical sound. We know from speaking voice research that emotions are reflected, for example, in voice quality.[1,2,3] For singers, the act of emotional expression is particularly demanding because the voice apparatus is already working at full capacity for singing alone.

The optimal function of the voice in singing is usually achieved through the physical activity of the body, the breathing mechanism, the laryngeal muscles, and the articulators, which we refer to as "the singing technique". The continuously changing optimal alignment of bony and cartilaginous structures along with exactly the right amount and distribution of muscle function are required to be able to achieve each pitch in a piece of music with a stylistically relevant timbre of voice.[4,5,6]

Emotions change the voice tone,[7] often deteriorating it from optimally balanced phonation; this is true for both the singing and the speaking voice. In the world of singing, there are genre-specific esthetic quality standards to which a singer must adjust. The stylistic differences also manifest themselves under the umbrella concepts of "classical" and "popular" music. There are distinct technical differences in singing baroque vs. opera or musical theatre vs. soul.[4] When expressing emotions, singers need to be aware of the effects that emotional expression exert on the voice so they can send their acoustic message without compromising the sonority of the voice too greatly.

1.1

Voice quality in singing

Much of the emotional information in the speaking voice is perceived from pitch, tempo, loudness, and rhythm, the use of which is restricted in musical performance.[8,9] If a singer follows the written music very strictly, the voice quality is really the only parameter that can be freely varied. This is true for both contemporary commercial music (CCM) and classical styles of singing. In speech and singing, the term "voice quality" refers to "the long-term auditory coloring of a

person's voice."[10] In a broad sense, it is the combined product of both laryngeal (phonation-related) and supralaryngeal (articulation-related) factors. [10]

Different styles of singing require the use of different voice qualities.[11,12,13] One needs to configure the phonation and articulatory settings differently for almost every musical style. For example, in bossa nova, the phonation settings are often breathy, whilst articulatory settings function at full throttle to make rhythmical distinctions. One of the key technical elements in opera is to be loud enough to be heard over the orchestra without electric sound amplification, so the singer configures the apparatus to take maximum advantage of vocal tract resonances.[14] In some styles of the heavy metal, singers need to adorn the voice with constant distortion, making the underlining voice quality sound harsh.[15] On top of this rather stable "stylistic voice quality," a singer makes another layer of smaller changes that mark the rendition of the emotional content of a song. Regardless of the genre or emotional content of the song, the singer needs basic skills to control the vocal apparatus to match pitches, produce dynamic variation, and deliver efficient articulation and various side sounds (like sighs, grunts, etc.) where needed.

1.2

Research questions

The recognition of emotion from the singing voice has been traditionally studied using short samples and listener group ratings.[16,17] Previous research has shown that in music, it is often the case that general categories of emotion are well recognized, but nuances (such as hot anger vs. cold anger) within these categories are not.[9] To our knowledge, the recognition of emotions between different styles of singing has not yet been studied. In the present paper, we study the recognition of four emotions (anger, sadness, tenderness, and joy). These emotions have been selected because they can be placed on a fourfold table of valence and activation. Anger, sadness, and joy are regarded as basic emotions and should by definition be easy to recognize.[18,19] Tenderness is included because an emotion with a positive valence and low activity level was needed to complete the fourfold table. All of these emotions occur frequently in song interpretations in both the classical and contemporary commercial worlds, and are thus familiar performance tasks for most singers.

In this preliminary study, we use short vowel samples as test material in order to investigate the role of voice quality in conveyance of emotions. Short vowels are used since they do not contain semantic or prosodic information. Therefore, they carry voice quality in its purest sense. Although the recognition percentage is not supposed to be high in short samples lacking prosodic information, earlier research on both speaking and singing voice suggests that emotional information can be received also from short samples[2,20,21]

The study is an experimental comparative design using singing voice samples and listener evaluations. The specific research questions of this study are:

1. Are listeners able to recognize emotions in a singing voice from short vowel samples?
2. Is there a difference in the recognition of emotions when they are sung using a classical singing technique compared to when they are sung using a CCM style of singing?
3. Does pitch affect the recognition of emotion in the classical-/CCM-style singing voice samples?
4. Are valence and the activation of the emotions perceptible in the sung samples?

2.

Methods

Thirteen professionally trained singers sang a small excerpt of a song expressing four different emotions. A listening test was created to determine whether the listeners' appraisals of the sung emotions matched the singers' intended expressions.

2.1

Participants and recording

The voice samples were gathered from 13 singers (two males, 11 females) with a minimum of three years of singing lessons at a professional level. The mean age of the subjects was 32 years (range: 20–44 years). The mean number of years' singing experience was 10. Singers were either classically trained (N=7) or CCM singers (N=6), and all of the subjects were native Finnish-speakers. Six of the singers worked in both classical and CCM genres, and these subjects gave voice samples in both styles.

The singers were instructed to perform an eight-bar excerpt from a song expressing the emotions of joy, tenderness, sadness, or anger using either a CCM or classical technique. The song was Gershwin's *Summertime* with Finnish lyrics by Sauvo Puhtila. This song was chosen because it has been composed as an aria, but it is widely popular among CCM singers as well, so it fits both the classical and the CCM repertoire. The Finnish lyrics depict a nature scene that contains no particular emotional information as such. An effort was made to make the experiment as lifelike as possible. Therefore, the singers used a backing track with a neutral accompaniment suitable for both classical and CCM style singing that was played to them via an S-LOGIC ULTRASONE Signature PRO headset. The studio setup also featured a Shure SM58 vocal microphone, which allowed the singer's singing voice to be mixed in with the backing track as they were singing.

Because pitch is known to vary in the expression of emotions in speech, and it could thus be expected to affect the perception of emotions, all subjects were instructed to use the same pitch (with the males singing one octave lower) regardless of genre.

The key of the song was D minor, and the tempo was 80 beats per minute for all test subjects and every emotional portrayal. The emotion samples were performed in a randomized order and repeated three times. The singers also gave a neutral voice sample without expressing any emotion. This sample was also repeated three times.

All recordings were made at the recording studio of Tampere University Speech and Voice Research Laboratory using a Brüel & Kjær Mediator 2238 microphone. The distance between the microphone and test subjects' lips was 40 cm. The samples were recorded with Sound Forge 7 digital audio editing software with a 44.1 kHz sampling rate using a 16-bit external soundcard (Quad-Capture Roland). All samples were saved as wav. files for further analysis with Praat.[22]

2.2 Voice samples

The vowel [ɑː] was extracted from three different pitches in each sample for further analyses. The pitches were, for the females, a, e1, a1 (A3, E4, A4 according to the American system), and A, e, and a for the males (A2, E3, A3). The [ɑː] samples were extracted from the Finnish words *aikaa* [ˈɑikɑː] (time), *hiljaa* [ˈçilˌjɑː] (softly), and *saa* [sɑː] (to be allowed). The nominal duration of the

extracted vowels (including the preceding consonant) were 2.25 s for a, 4.5 s for e1 and 2.25 s for a1, according to the notation and tempo of the song.

The [ɑ:] vowels were extracted from the sung excerpts using Sound Forge 7 audio editing software. The samples were cut right after the preceding consonant. The duration of the sample varied between 0.6267 s and 4.8063 s depending on how the test subject had interpreted the time value of the notation. The tail end of the vowel was left as the singer interpreted it (the nominal note durations 2.25 s or 4.5 s), as previous studies have indicated that micromanaging the durations of written notes is one way of expressing emotions in the singing voice. [23]

From a total number of 900 samples, 300 samples were chosen for the listening test (see Table 1).

Table 1: Numbers of voice samples in the listening test (total number of samples N=300).

	JOY		**TENDERNESS**		**SADNESS**		**ANGER**		**NEUTRAL**	
	Classical	CCM	Classical	CCM	Classical	CCM	Classical	CCM	Classical	CCM
High pitch	9	11	9	11	9	11	9	11	9	11
male 220Hz	1	1	1	1	1	1	1	1	1	1
female 440Hz	8	10	8	10	8	10	8	10	8	10
Medium pitch	9	11	9	11	9	11	9	11	9	11
male 165Hz	1	1	1	1	1	1	1	1	1	1
female 330Hz	8	10	8	10	8	10	8	10	8	10
Low pitch	9	11	9	11	9	11	9	11	9	11
male 110Hz	1	1	1	1	1	1	1	1	1	1
Female 220Hz	8	10	8	10	8	10	8	10	8	10

2.3 Listening test

The listening task was an internet-based test with 300 randomized [ɑ:] vowel samples and 12 control samples. As the samples were numerous, we constructed the listening test so that it was possible to stop and continue the test as needed. The test was accessible through a browser by logging in with a password. Participants completed the test using their own equipment. The voice samples were played in a randomized order and it was possible to play the samples as many times as needed. The Finnish questionnaire was translated for the one listener who was not

Finnish-speaking. We used Pearson's Chi-squared test of homogeneity to determine if the probability of recognition was the same for the native Finns and the non-native participant in order to check for the possible language-related or cultural differences in emotion recognition. The zero hypothesis that recognition percentage for group 1 (28 listeners) equals that of group 2 (1 listener) was to be accepted (*z*-value 2.0, *p*-value 0.160) at a statistical significance level of $\alpha = 0.05$. We also tested the possible effects of listening preferences to recognition by comparing the answers of participants who predominantly listen to classical singing with those who mostly listen to CCM. The zero hypothesis that recognition percentage for group 1 (those who mostly listen to CCM music, 26 listeners) equals that of group 2 (those who mainly listen to classical music, 3 listeners) was to be accepted (*z*-value 0.3, *p*-value 0.579) at a statistical significance level of $\alpha = 0.05$. The listening test took approximately 60 minutes to complete, and the participants were offered either study credits or voice lessons in exchange for the completed test. The number of people who completed the test was 29 (22 females, 7 males, no reported hearing defects), and they were all selected for further analysis. The listeners completed a multiple-choice questionnaire on which emotion they perceived (anger, sadness, joy, tenderness, neutral) for each sample. Eight of the listeners were professionally involved in assessing the human voice (singing teachers and vocologists) and 21 were laypeople. Seventeen of the listeners were singers (14 amateur and 3 professionals).

2.3.1 The number of samples used

The total number of samples listened by each listener was 312. There were 100 samples from each pitch, 60 samples + 3 control samples (repetitions of the same samples in a randomized order) for each emotion category, and 60 neutral samples. There were thus 20 samples depicting the same emotion category and pitch in the whole data sample. Of the samples, 135 were sung using a classical singing technique, and 165 samples were sung using a CCM-style singing technique.

The total number of answers in the listening test was 9,048. From each pitch, we gathered 3,016 answers. From each emotion category, we drew 1,827 answers, while 1,740 answers were drawn for the neutral portrayals. The samples where a classical style singing technique was used resulted in 3,915 answers and the samples where a CCM-style singing technique was used resulted in 5,133 answers.

Table 2: Numbers of answers given in the listening test (total number of samples N=9,048).

	JOY		TENDERNESS		SADNESS		ANGER		NEUTRAL	
	Classical	CCM	Classical	CCM	Classical	CCM	Classical	CCM	Classical	CCM
High pitch	261	348	261	348	261	348	261	348	261	319
male 220Hz	29	29	29	29	29	29	29	29	29	29
female 440Hz	232	319	232	319	232	319	232	319	232	290
Medium pitch	261	348	261	348	261	348	261	348	261	319
male 165Hz	29	29	29	29	29	29	29	29	29	29
female 330Hz	232	319	232	319	232	319	232	319	232	290
Low pitch	261	348	261	348	261	348	261	348	261	319
male 110Hz	29	29	29	29	29	29	29	29	29	29
Female 220Hz	232	319	232	319	232	319	232	319	232	290

2.4 Statistical analyses

The results of the listening test were coded numerically for statistical analysis. Both the intended and perceived emotions were given numbers (1= joy, 2 = tenderness, 3 = neutral, 4 = sadness, 5 = anger). The valence and activation of the emotions expressed and perceived were given arbitrary numbers based on the emotions chosen in the listening test. Positive valence (emotion that is regarded as pleasant) was coded as 2, negative valence as 1, and neutrality as 0. Activity (the energy level typically inherent in an emotion) was coded as low = 1, high = 2, or medium = 0. The pitch levels were coded as 1 = low, 2 = middle, or 3 = high pitch. The samples sung with a classical technique were marked as 1, and those sung in a CCM style were marked as 2.

The number of the correct (intended = perceived) answers for emotion, valence, and activity are given as percentages.

Furthermore, the results of the listening test were analyzed using three different statistical tests.

The first statistical test used was a binomial test (one-proportion *z*-test) to evaluate the probability that the observed percentage of the correctly recognized emotions could have resulted from random guessing. The listening test contained five different emotions, which meant that the

expected percentage of correctly recognized emotions in case of random guessing would be 20%. The observed percentage of correctly recognized emotions differs statistically significantly from random guessing if the *p*-value is <0.05.

The second statistical test, Pearson's Chi-squared test of homogeneity, was used to evaluate the probability that two groups of results have the same percentage of correctly recognized emotions. The percentage of correctly recognized emotions is statistically different in two groups of results if the *p*-value of the test is <0.05.

The third statistical test, Cronbach's alpha, was used to evaluate the reliability of the internal consistency of listener evaluations. Alpha values >0.7 indicate an acceptable internal consistency of the data.

Intra-rater reliability was estimated using Cohen's kappa coefficient.

All statistical analyses were performed using Microsoft Excel.

3.

Results

3.1

Accuracy of emotion recognition

The percentage of correct recognitions out of all the answers in the listening test (N=9,048) was 30.2%. According to the one-proportion *z*-test, the recognition exceeded random guessing (H0: *p*=1/5: *z*-value 24, *p*-value <<.001). The internal consistency of the listeners' evaluation was good (Cronbach's alpha 0.89). The intra-rater reliability was moderate (mean Cohen's kappa 0.48) according to the Landis and Koch benchmark.[24] In females, the emotions were recognized significantly better from the CCM samples than from the classical samples. Recognition from the female samples sung using a CCM style was higher (1,653 correct answers from 4,698 answers given) than from the samples sung in a classical style (832 correct answers from the total number of 3,480 answers given). The statistical significance of the 11.3% difference in recognition was

evaluated using the Pearson's Chi-square test of homogeneity, and the difference was found to be significant (*z*-value 120.2, *p*-value<<.001). Recognition from the male singers' CCM- and classical-style samples was not statistically significantly different (Pearson's Chi-square *z*-value 0.5 and *p*-value 0.5). Correct recognition occurred in 119 answers from a total of 435 given for a CCM style and 128 answers from a total of 435 given for the classical style.

The discrepancy between the number of CCM and classical female samples was considered a possible factor in information distortion. We performed Pearson's Chi-squared test of homogeneity to test if the probability of recognition was statistically the same for these two groups. To validate the use of a larger sample number in one group, we excluded the two best recognized female CCM singers from the sample battery, and compared the correctly recognized samples of the nine least well recognized female CCM singers with the female classical singers. Pearson's Chi-squared test of homogeneity indicated that on the statistical significance level $\alpha \leq 0.05$, the zero hypothesis that the recognition percentage for group 1 (11 CCM singers) equals that of group 2 (nine classical singers) had to be discarded. The same was true when comparing group 1 (the nine least well recognized CCM singers) with group 2 (nine classical singers). Thus, even after excluding the two best recognized sample batteries among the female CCM samples, the recognition of the CCM samples remained significantly better than that of the classical samples. Therefore, the discrepancy between the number of CCM and classical samples has not affected the results.

There was a median 3.9% difference in the overall recognition of emotions from the low frequency to the high frequency in such way that the low frequency samples were systematically recognized more poorly than the high frequency samples in all sample groups (see Table 3).

Table 3: Correctly recognized samples, differences in recognition between CCM styles of singing and classical singing at three different pitches, and the internal consistency of the answers (statistical significance level $\alpha \leq 0.05$).

			%	***z*-value**	***p*-value**	**Cronbach's alpha**
Female	**CCM**	**Overall recognition**	**35.2%**	**26.02**	**0.000**	**0.90**
		low pitch	34.7%	14.51	0.000	0.92
		medium pitch	35.1%	14.90	0.000	0.88
		high pitch	35.9%	15.72	0.000	0.90
	Classical	**Overall recognition**	**23.9%**	**5.76**	**0.000**	**0.88**
		low pitch	22.3%	1.94	0.052	0.93
		medium pitch	24.2%	3.60	0.000	0.85

		high pitch	25.3%	4.48	0.000	0.82
Male	**CCM**	**Overall recognition**	**27.4%**	**3.84**	**0.000**	**0.85**
		low pitch	22.1%	0.62	0.533	0.87
		medium pitch	26.9%	2.08	0.000	0.85
		high pitch	33.1%	3.94	0.000	NaN
	Classical	**Overall recognition**	**29.4%**	**4.91**	**0.000**	**0.80**
		low pitch	26.9%	2.08	0.038	0.93
		medium pitch	29.7%	2.91	0.004	0.78
		high pitch	31.7%	3.53	0.000	NaN

Table 4 indicates that in the case of all other emotions except sadness, recognition of the emotional content from the sung [a:] vowels from the female singers was easier when the samples were sung using a CCM-style technique. Sadness, on the other hand, was better recognized from the samples sung with a classical-style technique. From the male singers' samples, joy, tenderness, and neutral portrayals were recognized more accurately from the CCM samples, whilst sadness and anger were recognized more accurately from the classical singing technique. The correct perception of anger seemed to be clearly easier from female CCM samples than from any other samples.

Table 4: Correctly recognized emotions, differences in recognition between CCM and classical singing, and the internal consistency of the answers (statistical significance level $\alpha \leq 0.05$).

			%	*z*-value	*p*-value	Cronbach's alpha
Female	**CCM**	**Joy**	24.3%	3.36	0.001	0.90
		Tenderness	33.1%	10.15	0.000	0.78
		Neutral	29.5%	7.03	0.000	0.73
		Sadness	34.5%	11.2	0.000	0.89
		Anger	53.9%	26.23	0.000	0.95
	Classical	**Joy**	14.5%	-3.62	0.000	0.89
		Tenderness	13.4%	-4.38	0.000	0.56
		Neutral	28.7%	5.76	0.000	0.39
		Sadness	36.2%	10.69	0.000	0.86
		Anger	26.7%	4.43	0.000	0.95
Male	**CCM**	**Joy**	12.6%	-1.72	0.086	0.73
		Tenderness	31%	2.57	0.01	0.78
		Neutral	40.2%	4.75	0.000	0.66
		Sadness	34.5%	3.38	0.000	0.81
		Anger	18.4%	-0.38	0.707	0.97
	Classical	**Joy**	11.5%	-1.98	0.047	0.91
		Tenderness	27.6%	1.77	0.077	0.79
		Neutral	36.8%	3.91	0.000	-0.69
		Sadness	50.6%	7.13	0.000	0.76
		Anger	20.7%	0.16	0.872	-0.09

Pitch played a role in emotion recognition. Sadness was more easily recognized from a low pitch (female voice 55.9%, male voice 55.2%) and less easily recognized from a high pitch (females 15.6%, males 24.1%). The recognition of joy was better from a high pitch (females 42%, males 28.1%) and more poorly from a low pitch (females 5.4%, males 0%). The recognition of tenderness was slightly better at a middle pitch (females 31.6%, males 41.4%). Anger was best recognized from high frequencies in all sample groups. (See Table 5.)

We tested the internal consistency of the answers in the female samples at different pitches with Cronbach's alpha and it showed a mean consistency of 0.60. However, the fluctuation of listener agreement between emotions was considerable (Cronbach's -0.37–0.97). Anger yielded the most consistent answers, while joy yielded the least consistent answers.

When comparing the recognition of emotions from different pitches in the classical and CCM styles of singing, the most prominent difference was seen in the recognition of anger in the female

classical and male CCM samples. Anger was perceived 26.6% units better at a high pitch than at a low pitch from the female classical samples and 55.2% units better at a high pitch than at a low pitch from the male CCM samples. (See Table 5).

3.2

Valence and activation appraisals

Valence and activation were derived from the listeners' answers. Of the samples produced by female CCM singers, valence was correctly perceived as positive 887 times (46.3%) and as negative 997 times (52.1%). From the female classical samples, valence was correctly perceived as positive 402 times (28.9%) and as negative 698 times (50.1%). From the male CCM samples, valence was correctly perceived as positive 63 times (36.4%) and as negative 63 times (36.2%). From male classical samples, valence was correctly perceived as positive 57 times (32.8%) and as negative 75 times (43.1%).

Of the samples produced by female CCM singers, activation was correctly perceived as high 967 times (50.5%) and as low 1145 times (59.8%). From the female classical samples, activation was correctly perceived as high 583 times (41.9%) and as low 733 times (52.7%). From the male CCM samples, activation was correctly perceived as high 48 times (27.7%) and as low 102 times (58.6%). From the male classical samples, activation was correctly perceived as high 47 times (27%) and as low 120 times (69%).

In the answers given, valence was perceived with a 41.6% accuracy, and activity with a 45.8% accuracy. High activity was perceived with a 41.5% accuracy and low activity with a 57.5% accuracy. Positive valence was perceived with a 38.6% accuracy and negative valence with a 50.2% accuracy.

When comparing the perceived accuracy of valence and activation between the CCM style and classical style, we can see that in the female samples, both valence and activation are more accurately perceived from the CCM-style samples, where, as with the male samples, they were better perceived compared to the classical-style samples. (See Table 5.)

In these data, activation was more accurately perceived from all pitches in comparison to valence. At a low pitch, valence was more accurately perceived for joy and anger, whereas activation was more accurately perceived for tenderness and sadness. At a middle pitch, the tendency was similar with the female samples, but with the male samples, the valence was more accurately perceived only for joy. At a high pitch, activity was perceived more accurately for all other emotions except for female tenderness, in which valence was

more accurately perceived. The perception of valence and activation from the female samples was most accurate at a high pitch. From the male samples, valence was correctly perceived from a middle pitch most accurately, whereas activation was most accurately perceived from a high pitch.

Table 5: Correct recognition of emotion, valence, and activation at different pitches.

	Females		Males	
	CCM	**Classical**	**CCM**	**Classical**
Low pitch (220Hz/ 110Hz)				
JOY	8.20%	1.70%	0.00%	0.00%
Valence	29%	9.50%	17%	14%
Activation	17.20%	16.80%	7%	0%
TENDERNESS	24.50%	9.50%	37.90%	10.30%
Valence	26.30%	11.20%	38%	14%
Activation	68.00%	65.50%	72.40%	69%
SADNESS	58.00%	53.00%	41.00%	69.00%
Valence	59.90%	60.30%	44.80%	59%
Activation	70.50%	63.40%	69.00%	86%
ANGER	49.80%	15.60%	0.00%	13.80%
Valence	69.00%	63.40%	41%	31.00%
Activation	53.90%	17%	3%	24.10%
Middle pitch (330Hz/165Hz)				
JOY	16.60%	8.20%	13.80%	3.40%
Valence	40.80%	30.60%	65.50%	41.40%
Activation	30.70%	14.20%	14%	3%
TENDERNESS	39.50%	20.70%	41.40%	41.40%
Valence	52.40%	27.20%	41%	48.30%
Activation	64.90%	66.40%	75.90%	69%
SADNESS	31%	38%	48%	48%
Valence	32.60%	40.10%	48%	51.70%
Activation	64.30%	61.20%	82.20%	65.50%
ANGER	56.10%	22.40%	0%	20.70%
Valence	65.80%	48.30%	10%	28%
Activation	64.30%	30.20%	14%	35%
High pitch (440Hz/220Hz)				
JOY	48.30%	33.60%	24.10%	31%
Valence	58.10%	40.90%	32.10%	37.90%
Activation	63.80%	56.50%	60.70%	37.90%
TENDERNESS	35.40%	9.90%	13.80%	31%
Valence	71.20%	53.90%	24.10%	41%
Activation	48.60%	28.90%	41.40%	58.60%
SADNESS	14%	17%	14%	35%
Valence	18.20%	30.20%	17%	37.90%

	Activation	42.60%	30.60%	20.70%	65.50%
ANGER		56.10%	42.20%	56%	27.60%
	Valence	67.10%	58.60%	55%	41.40%
	Activation	73.00%	61%	69%	62.10%

4.

Discussion

The percentage of correctly recognized emotion samples in this study was relatively low (30.2%) compared to earlier studies concerning speech. Most of the studies examining emotion recognition from the speaking voice reach recognition percentages above 50%.[1,25,26,20,27]Thus, it seems to be harder to recognize emotions from singing samples, at least when they are short.

Previous research suggests that the expression of emotions in the singing and speaking voice are related,[28] and that the same methods of emotion recognition apply to both.[29] As with the speaking voice, the voice quality in anger is easiest to recognize. This phenomenon might have an evolutionary underpinning, as it continues to be a useful skill to recognize potentially hazardous situations.

In this data, emotional content from the CCM-style samples was correctly recognized 11.3% more often than from the classical-style samples. The recognition of sadness, however, was 3.3 percent units higher from the samples sung using the classical techniques. It could be postulated that the darker timbre typical in classical singing makes it easier to be interpreted as related to sadness. The dark timbre in classical singing is due to lower resonance frequencies related to the lowering of the larynx and, thus, the lengthening of the vocal tract.[30] According to earlier studies, a strong fundamental, relatively weak overtones near 3kHz, and lack of vibrato have been found to be indicative of typical expressions of sadness in the Western classical singing style.[17,31,32] However, many samples that were recognized as expressions of sadness in the present study had a very clear vibrato. Another possible explanation for the results of the present investigation (made in Finland) could be cultural. The Russian lament uses a simultaneous amplitude and frequency modulation that is reminiscent of vibrato. Another factor could be the distinct distribution of energy during one vowel, where intensity increases towards the end: this technique is also used in

laments.[33] Further investigation is needed to examine the acoustic structure of the voice samples in this study.

The speech-like qualities of the CCM style of singing are one possible explanation as to why it seems to be easier to recognize emotion portrayals from it. Another possibility as to why the CCM style of singing is more recognizable could be that it uses the 'chest voice' more often, whereas classical singing operates more with the 'head voice'. In the chest voice, the mass of the vocal folds vibrates more vertically, making a more robust impact on air pressure, and the formants appear easier.[34] The slope of the sound spectrum is more gradual, as the relative amplitude of the upper partials is more pronounced. In the head voice, the slope is steeper.[35] It is also plausible that speakers and CCM-style singers use more variation in phonation type along the axis from breathy to pressed,[36] while classical singers keep the voice source more stable. This is related to both esthetic and technical demands. For instance, pressed phonation may make it more difficult or even impossible to reach the high pitches required.

Another possible explanation for the better identification of emotion in CCM-style singing is familiarity. In general, most people are exposed to far more CCM than classical singing. Therefore, they may be more attuned to emotion in these genres.

Previous research has shown that there is a considerable variability in the individual ability to express emotion by singing.[16] Mirroring this fact, the two male singers who participated in this study were too few to properly represent male CCM and classical singers. However, we wanted to keep them in the study because despite their individuality, the listening test appraisals were mostly very similar to those given for the female samples. The individual ability to express emotions was also seen in the female samples. When we excluded the two best recognized CCM female singers from the sample size, the groups of all CCM singers (N=11) and CCM singers -2 (N=9) were no longer statistically a part of the same group (Pearson's Chi-squared test of homogeneity). This suggests that in future studies, the sample size should be fairly large in order to increase validity.

Pitch seemed to affect the assessment of emotions. As in the speaking voice, the higher the pitch, the more often the listeners chose an emotion that represents a high overall activity level (Tables 5–6). This is understandable, since a higher pitch is typically produced with higher subglottic pressure and thus intensity.[6,35] This phenomenon is potentially counter-productive for singers

needing to portray non-active negative emotions (like sorrow) at a high frequency or high-activation positive emotions (like joy) at a low frequency. The tendency not to recognize joy but to recognize sadness was very pronounced at a low pitch (220/110Hz) for both the female and male samples. At a high pitch (440Hz), the phenomenon was reversed.

The pitches were selected with the female singers in mind from a pitch range that would allow singing the whole eight-bar song in either the chest or the head register, should the singer choose to express it so. This was done to accommodate various singing styles and make as much room as possible for emotional expression whilst gaining data from the same pitches. It is possible that the choice to use the same pitches for all subjects and both singing styles may have somewhat interfered with the results, since the pitch range was somewhat low for classical singing. On the other hand, the participants were trained singers who ought to be well able to sing at these pitches.

The accuracy of perceived valence and activity in the listening test answers may suggest that it is easier to make assessments of valence and activity than to recognize emotions *per se*. This corresponds to the earlier findings for speech. Similarly, we found in the female samples that samples with a negative valence and high activity were more easily recognized than those with a positive valence and low activity. This is understandable, since it is important for survival to be able to quickly recognize signs of potentially dangerous situations.

5.

Conclusions

1) Emotions were recognized above the level of chance in short vowel extracts from singing.
2) Emotions were recognized statistically significantly better in the samples with a CCM style of singing compared to the samples featuring classical singing.
3) Pitch also plays a role in emotion recognition in the singing voice.
4) The valence and activation levels of the voice also play a role in emotion recognition in singing.

Acknowledgments
This research was supported by the Eemil Aaltonen Foundation through a grant (160036 N1) and

by the Oskar Öflunds Stiftelse Foundation through a grant to Tua Hakanpää. The authors would like to thank Antti Poteri, D.Sc. (Tech), for help with the statistical analysis.

References

1. Banse R, Scherer KR. Acoustic profiles in vocal emotion expression. *J Pers Soc Psychol*. 1996;70(3):614-636. doi:10.1037/0022-3514.70.3.614.

2. Laukkanen A-M, Vilkman E, Alku P, Oksanen H. On the perception of emotions in speech: the role of voice quality. *Logop Phonatrics Vocology*. 1997;22(4):157-168. doi:10.3109/14015439709075330.

3. Tartter VC. Happy talk: Perceptual and acoustic effects of smiling on speech. *Percept Psychophys*. 1980;27(1):24-27. doi:10.3758/BF03199901.

4. Hallqvist H, La FMB, Sundberg J. Soul and Musical Theater: A Comparison of Two Vocal Styles. *J Voice*. 2016. doi:10.1016/j.jvoice.2016.05.020.

5. Björkner E. Musical Theater and Opera Singing-Why So Different? A Study of Subglottal Pressure, Voice Source, and Formant Frequency Characteristics. *J Voice*. 2008;22(5):533-540. doi:10.1016/j.jvoice.2006.12.007.

6. Titze IR. *Principles of Voice Production*. Englewood Cliffs, NJ: Prentice-Hall; 1994.

7. Williams CE, Stevens KN. Emotions and speech: some acoustical correlates. *J Acoust Soc Am*. 1972;52(4):1238-1250. doi:10.1121/1.1913238.

8. Chua G, Chang QC, Park YW, Chan PY, Dong M, Li H. The expression of singing emotion - Contradicting the constraints of song. In: *Proceedings of 2015 International Conference on Asian Language Processing, IALP 2015*. ; 2016:98-102. doi:10.1109/IALP.2015.7451541.

9. Juslin PN, Laukka P. Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol Bull*. 2003;129(5):770-814. doi:10.1037/0033-2909.129.5.770.

10. Laver J. the Phonetic Description of Voice Quality. *New York*. 1980;31:66-92. http://cat.inist.fr/?aModele=afficheN&cpsidt=12583168.

11. Manfredi C, Barbagallo D, Baracca G, Orlandi S, Bandini A, Dejonckere PH. Automatic Assessment of Acoustic Parameters of the Singing Voice: Application to Professional Western Operatic and Jazz Singers. *J Voice*. 2015;29(4):517.e1-517.e9. doi:10.1016/j.jvoice.2014.09.014.

12. Schloneger MJ, Hunter EJ. Assessments of voice use and voice quality among college/university singing students ages 18-24 through ambulatory monitoring with a full accelerometer signal. *J Voice*. 2017;31(1):124.e21-124.e30.

13. Barlow C, LoVetri J. Closed quotient and spectral measures of female adolescent singers in different singing styles. *J Voice*. 2010;24(3):314-318.

14. Sundberg J, Lã FMB, Gill BP. Formant tuning strategies in professional male opera singers. *J Voice*. 2013;27(3):278-288. doi:10.1016/j.jvoice.2012.12.002.

15. Borch DZ, Sundberg J, Lindestad P a, Thalén M. Vocal fold vibration and voice source

aperiodicity in “dist” tones: a study of a timbral ornament in rock singing. *Logoped Phoniatr Vocol*. 2004;29(6):147-153. doi:10.1080/14015430410016073.

16. Siegwart H, Scherer KR. Acoustic concomitants of emotional expression in operatic singing: The case of lucia in Ardi gli incensi. *J Voice*. 1995;9(3):249-260. doi:10.1016/S0892-1997(05)80232-2.

17. Scherer KR. Expression of emotion in voice and music. *J Voice*. 1995;9(3):235-248. doi:10.1016/S0892-1997(05)80231-0.

18. Ekman P. Are there basic emotions? *Psychol Rev*. 1992;99(3):550-553. doi:10.1037/0033-295X.99.3.550.

19. Izard CE. Basic emotions, relations among emotions, and emotion-cognition relations. *Psychol Rev*. 1992;99(3):561-565. doi:10.1037//0033-295X.99.3.561.

20. Waaramaa T, Laukkanen AM, Alku P, Väyrynen E. Monopitched expression of emotions in different vowels. *Folia Phoniatr Logop*. 2008;60(5):249-255. doi:10.1159/000151762.

21. Airas M, Alku P. Emotions in vowel segments of continuous speech: Analysis of the glottal flow using the normalised amplitude quotient. *Phonetica*. 2006;63(1):26-46. doi:10.1159/000091405.

22. Boersma P, Weenink D. Praat. 2014.

23. Sundberg J. Emotive Transforms: acoustic patterning of speech Its Linguistic and Physiological Bases. *Phonetica*. 2000;(57):95-112.

24. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174. doi:10.2307/2529310.

25. Scherer KR. Vocal communication of emotion: A review of research paradigms. *Speech Commun*. 2003;40(1-2):227-256. doi:10.1016/S0167-6393(02)00084-5.

26. Scherer KR, Banse R, Wallbott HG. Emotion Inferences from Vocal Expression Correlate Across Languages and Cultures. *J Cross Cult Psychol*. 2001;32(1):76-92. doi:10.1177/0022022101032001009.

27. Iida A, Campbell N, Higuchi F, Yasumura M. A corpus-based speech synthesis system with emotion. *Speech Commun*. 2003;40(1-2):161-187. doi:10.1016/S0167-6393(02)00081-X.

28. Scherer KR, Sundberg J, Tamarit L, Salom??o GL. Comparing the acoustic expression of emotion in the speaking and the singing voice. *Comput Speech Lang*. 2015;29(1):218-235. doi:10.1016/j.csl.2013.10.002.

29. Eyben F, Salomão GL, Sundberg J, Scherer KR, Schuller BW. Emotion in the singing voice—a deeperlook at acoustic features in the light ofautomatic classification. *EURASIP J Audio, Speech, Music Process*. 2015;2015(1):19. doi:10.1186/s13636-015-0057-6.

30. Miller Richard. *The Structure of Singing, System and Art in Vocal Technique*. Belmont CA: Wadsworth Group; 1996.

31. Sundberg J. Expressivity in singing. A review of some recent investigations. *Logop Phoniatr Vocology*. 1998;23(3):121-127. doi:10.1080/140154398434130.

32. Jansens S, Bloothooft G, de Krom G. Perception And Acoustics Of Emotions In Singing. *Proc Fifth Eur Conf Speech Commun Technol*. 1997;0:0-3. http://citeseerx.ist.psu.edu/viewdoc/summary;jsessionid=9747D0A838F2790BD0161DCF94739C2E?doi=10.1.1.56.8871.

33. Mazo M, Erickson D, Harvey T. Emotion and Expression: Temporal Data on Voice Quality in Russian Lament. In: *Vocal Fold Physiology Voice Quality Control*. ; 1995:173-187.

34. Titze IR, Martin DW. Principles of Voice Production. *J Acoust Soc Am*. 1998;104(3):1148. doi:10.1121/1.424266.

35. Sundberg J. *The Science of the Singing Voice*.; 1987.

36. Peterson KL, Verdolini-Marston K, Barkmeier JM, Hoffman HT. Comparison of aerodynamic and electroglottographic parameters in evaluating clinically relevant voicing patterns. *Ann Otol Rhinol Laryngol*. 1994;103(5 Pt 1):335-346. doi:10.1177/000348949410300501.